**Data Science Interview Questions - Python and Tools**

*Source: Interview Preparation*

Python and Data Science Tools Interview Questions:  
   
 1. How do you handle large datasets in pandas?  
 Techniques: chunking, dask, vaex, using appropriate dtypes,   
 memory optimization, using SQL databases for very large datasets.  
   
 2. What's the difference between merge and join in pandas?  
 merge() is more flexible and can handle different join types.   
 join() is a convenience method for joining on index.  
   
 3. How do you optimize pandas performance?  
 Use vectorized operations, avoid apply() when possible, use appropriate dtypes,   
 use numba for custom functions, use swifter for parallel apply.  
   
 4. Explain the difference between loc and iloc in pandas.  
 loc uses labels/boolean indexing, iloc uses integer positions.   
 loc is inclusive of the end label, iloc is exclusive.  
   
 5. How do you handle categorical variables in machine learning?  
 Methods: one-hot encoding, label encoding, target encoding,   
 feature hashing, embedding layers for deep learning.  
   
 6. What's the difference between fit() and fit\_transform() in sklearn?  
 fit() learns parameters from data, fit\_transform() learns and transforms in one step.   
 transform() applies learned parameters to new data.  
   
 7. How do you implement cross-validation in sklearn?  
 Use cross\_val\_score(), KFold, StratifiedKFold, or custom CV splits.   
 For time series, use TimeSeriesSplit.  
   
 8. How do you handle text data in machine learning?  
 Preprocessing: tokenization, stemming/lemmatization, stop word removal.   
 Vectorization: TF-IDF, word embeddings, BOW, n-grams.  
   
 9. What's the difference between train\_test\_split and cross-validation?  
 train\_test\_split creates a single split, cross-validation creates multiple splits   
 for more robust evaluation.  
   
 10. How do you save and load machine learning models?  
 Use pickle, joblib (recommended for sklearn), or model-specific methods.   
 Consider model versioning and deployment requirements.